**ST**
Science & Technology

**PAPER – OPEN ACCESS**

# Dictionary Of Psychology With Horspool Algorithm

# TALENTA Conference Series

**S T**
Science & Technology

# Dictionary Of Psychology With Horspool Algorithm

Ade Mutiara Kartika[a*], Poltak Sihombing[a], Ade Candra[a]

*Computer Science Study Programme, Faculty of Computer Science and Information Technology, University of Sumatera Utara-20155 , Indonesia*

ademutiaraa@students.usu.ac.id, poltak@usu.ac.id, ade candra@usu.ac.id

**Abstract**

Dictionary is a reference that contains words and phrases that are usually sorted alphabetically and is followed by a description of the word or phrase. Dictionary consists of several types, one of which is a dictionary of terms. Psychology dictionary is a dictionary of terms that contain psychological term. Psychology dictionary is identical to the old tome in search and less efficient in its use. Made an Android-based application of psychology dictionary that applies the Horspool algorithm for terms search and SQLite for data storage is a solution. The search algorithm is needed to shorten the time in the search itself. Horspool algorithm apply a saved search information to shift even further because the string matching is done from right to left, so the time to search a string is shorter. Fields used in this research is psychology in general terms. By entering a psychological term as input in the search then we will get the search results. The results of this research show the entire pattern that matches the text.

*Keywords: Dictionary of Psychology; Horspool algorithm, String; Android.*

## 1. Introduction

### 1.1. Background

In the current era of globalization of information has grown rapidly. Smartphone as part of the globalization era, more developed and attractived its use by the society because of various features can be shown to fulfill the requirement and the attraction for users. The types of its operating systems including Windows, Blackberry, Symbian, iOS and Android. The Android operating system is one of the operating systems being developed in the community. There are advantages of this operating system, among others, the operating system can be modified in accordance with the wishes of our own according to Parno, et al [9].

The social demand for IT-based technology services very varied, one needs is the need for the availability of the dictionary with a variety of interests ranging from general dictionaries as dictionary to dictionary of specific terms such as political dictionaries, economic dictionaries and dictionary of psychology according to Parno, et al [9].

As you know, the science of psychology has foreign terms are not known by the layman. Dictionary of psychology currently circulating in the form of books, that its use is difficult because users need to find meaning and psychological term manually. On the other hand, the book is very difficult to carry everywhere and can not be updated at any time according to the needs and development at this time. Given the present-day society is very familiar with smartphone services, then the authors intend to make an application on the Android platform containing psychology terms using Horspool algorithm.

A string matching algorithm itself is the most important algorithms in text processing. Charras, C. et al [2] mentioned that this algorithm is also a basic component in the implementation of software tools in most current

operating systems at this time. One of a string matching algorithm is Horspool algorithm. Horspool algorithm is derived from the Boyer-Moore algorithm and easy in implementation. When the length of the pattern is small, highly inefficient to use the Boyer-Moore algorithm. Horspool algorithm uses only the bad-character displacement that occurs in the Boyer-Moore. To make and calculate the bad-character's shift value is to look at the right-most characters in the window. This shift value is calculated at the preprocessing stage for all the characters in the alphabet set previously. This algorithm is more efficient to use when the length of the small pattern is found by Sheik, S.S. et al [11].

### 1.2. Objective Of Research

The purpose of this study is to provide an application that can store and search for the psychology terms on android-based smartphone.

### 1.3. Construction of references

The expected benefits of this research is to help people in finding a psychological term that wish to know is much easier and more practical.

### 1.4. Methodology of Research

This research applies several research methods as follows:
- Study Literature, at this stage carried out a review of books, journals, articles that discuss Horspool algorithm.
- Analysis System, conduct an analysis of the existing problems, limitations of the system, the performance of the system, how the system works.
- System Design, make a flowchart, Unified Modeling Language (UML) and Interface Design.
- Implementation System, at this stage, the implementation Horspool algorithm to search for psychology terms in a dictionary application based android.
- Testing System, applications that have been implemented are tested based on the results of the analysis and design and test the correctness of the Horspool algorithm to find translation result data.
- Documentation System, documentation generated by making the thesis as a report of the research results.

## 2. Literature Review

### 2.1. Information Retrieval System

Information retrieval system is basically a process to identify, then call (retrieval) a document from a store (file), in response to a request for information by Hasugian, J [6]. Simply put, the information retrieval is a system that stores information and found the back of such information by Saptari, J. et al [10]. Information retrieval (Information Retrieval System) is the science of searching for information within a document, search for the document itself and look for metadata that describe a document. In point of view of the user, information retrieval help to search for information by providing a collection of information according to user needs by Wibowo, A [13]. There are five components of information retrieval according to Hasugian, J [6], namely:
- Users, are people who use or exploit IRS within the framework of management and information retrieval.
- Query, is a formatting language used by the user into IRS.
- Documents, are the term used to mention the entire material library e-book and e-journal.
- Document index, a glossary of terms or words (list of terms) are stored into a database that serves as a representation of a document.
- Matching, which is the term (query) entered by the user to index documents stored in the database is done by the engine computer.

*2.2. String Matching*

String matching is searching a pattern in a text by Cormen, T.H. et al [3]. Charras, C. et al [2] mentioned string matching is used to find a string called a pattern in a string called the text. As Effendi, D. et al [4] said that the working principles of string matching algorithm are as follows:

- Scan text with the help of a window the same size as the pattern length.

- Put the window at the beginning of the text.

- Compare the characters in the window with the character of the pattern. String matching algorithm has three main components, namely:

- Pattern, ie. sequences of characters that will be matched to the text, represented with x [0 . . . m − 1], the length of the pattern revealed by m.

- Text, where the pattern matching is done. Represented withy [0 n − 1], the length of the text revealed by n.

- Alphabet, contains all the symbols used by the language of the text and the pattern, expressed by Σ with ASIZE stated size.

*2.3. Horspool Algorithm*

Horspool algorithm is a simplification of the Boyer-Moore algorithm created by Nigel R. Horspool. The problems in the text searching is searching in large text to find the first pattern. Because the searched text can be quite large (allowing hundreds of thousands of characters) it is important to use a more efficient technique. Horspool algorithm works with a method similar to the Boyer-Moore algorithm but do not make the leap based on characters in the pattern that was found does not match the text by Horspool, R.N. [7].

Horspool algorithm has the right-most characters shift value of the window. At the stage of beginning observations (preprocessing), the shift value will be calculated for all the characters. At this stage, compared pattern from right to left until a match or mismatch pattern occurs. Right-most characters in the window is used as an index in making the shift value. In the case of incompatibility (characters not contained in the pattern) occurs, the window is shifted by the length of a pattern. Otherwise, Baeza-Yates, R.A. et al [1] mentioned that the window is shifted according right-most characters in the pattern.

There are two stages in the string matching algorithm using Horspool according to Singh, R. et al [12], which are:
- Preprocessing stage

At this stage, were observed pattern of the text to construct a table of bad-match that contains the value of the shift when the incompatibility between the pattern and the text occurs. Systematically, the steps undertaken at the stage preprocessing Horspool algorithm is:

- Horspool algorithms do the matching characters including the right pattern.
- Each character in the pattern added to the bad-match tables and calculated the value of its shift.
- Characters that are at the end of the pattern are not counted and not be a pitch-right character of the same character with him.
- When two characters are the same and one is not right-most characters, then characters with the largest index calculated the value of its shift.
- Horspool algorithm stores the length of the shift pattern as long as default values when a character in the text is not found in the pattern.

- Shift value to be used can be found by calculating the length of the pattern minus the last character index minus 1, for each character, $value = m − i − 1$.

- **Searching stage**

Systematically, the steps are performed on stage Horspool preprocessing algorithm is:
- Do comparison of the right-most character pattern to the window.
- Bad-match table used to skip characters when a mismatch occurs.
- When there is a mismatch, then the right-most character in the window serves as the foundation for determining the distance of the shift to be performed.
- After performing the matching (either the results is fit or not fit) do a shift to the right in the window.
- This procedure is repeated until the window is located at the end of the text or when the pattern is matched to the text.

To illustrate the details of the algorithm, will be given an example of a case where the pattern P = "Kartika" and the text T = "ADEMUTIARA KARTIKA". Initialization and manufacture of bad-match seen in Table 1 below.

Table 1: Preparation of bad-match

| P | A | R | T | I | K | * |
|---|---|---|---|---|---|---|
| i | 1 | 2 | 3 | 4 | 5 | - |
| v | 5 | 4 | 3 | 2 | 1 | 7 |

Table 1 as seen above shows the value of bad character shift-match by calculating the value of v. If there is a mismatch, there will be a shift to the right to skip characters that do not match where the value of the shift contained in the bad-match table, with the most-right characters as a reference. This can be seen in Table 2 below.

Table 2: Horspool algorithm iteration

| M | 1 | 2 | 3 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| T | A | D | E | M | U | T | I | A | R | A | | K | A | R | T | I | K | A |
| P | K | A | R | T | I | K | A | | | | | | | | | | | |
| I | 0 | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | | | | | |

`

Phase 1

| P | | K | A | R | T | I | K | A |
|---|---|---|---|---|---|---|---|---|
| I | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

Phase 2

| P | | | | K | A | R | T | I | K | A |
|---|---|---|---|---|---|---|---|---|---|---|
| I | | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

Phase 3

| P | | | | K | A | R | T | I | K | A |
|---|---|---|---|---|---|---|---|---|---|---|
| I | | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

In phase 3, as seen in Table 2, the window was located at the end of the text and all pattern matches with the text. The whole character matching using Horspool algorithm has been completed and stopped in the third phase.

## 3. Analysis And Design

### 3.1. Analysis of the Problem

To identify the described problem above is used Fishbone Diagram. Fishbone Diagram is a visual tool to identify, explore and describe all diagrams in a cause related to a problem. Fishbone Diagram can be
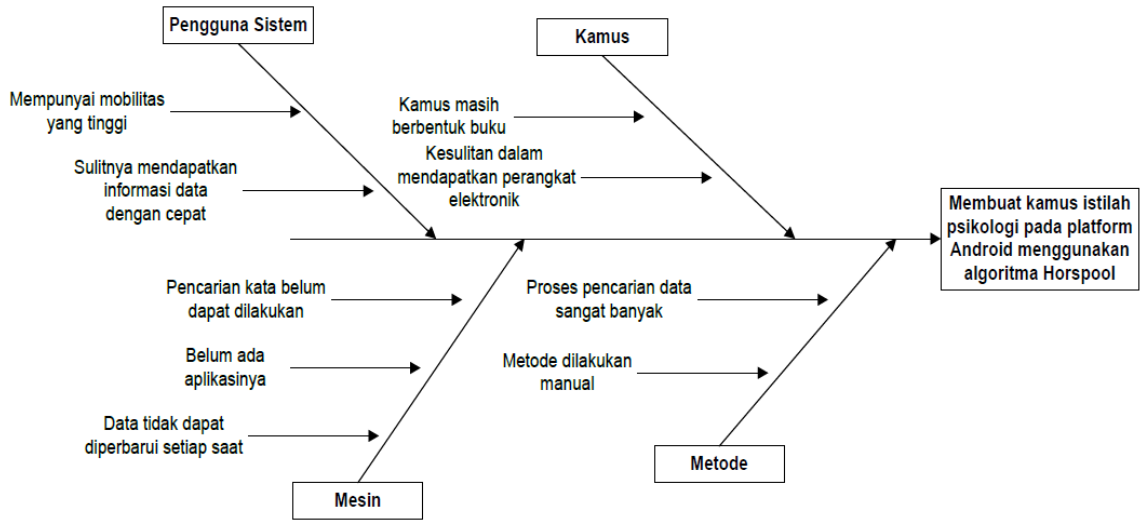


figure 1: Problem Analysis Fishbone Diagram

### 3.2. Modeling System

System modeling performed to obtain a clearer picture of any object that will interact with the system, as well as what things should be done by a system so that the system can function properly in accordance with its usefulness.

In this study used UML (Unified Modeling Language) as a modeling language for designing a system that will be built. UML that is used, among others, use case diagrams, activity diagrams and sequence diagrams.
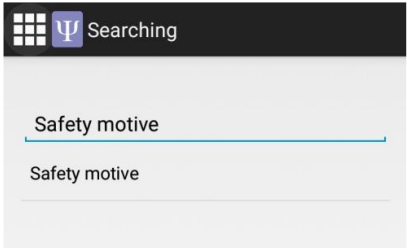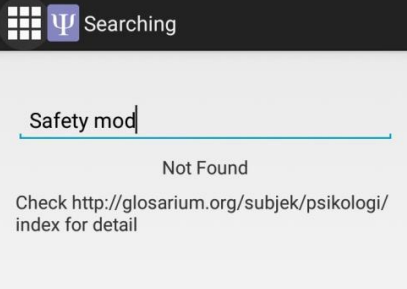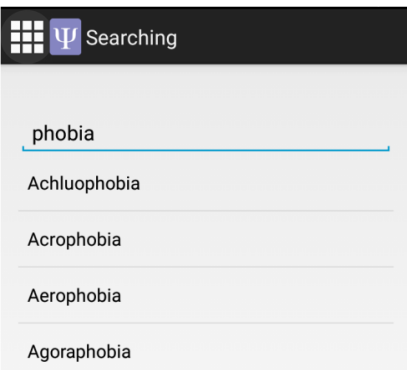
## 4. Implementation And System Testing
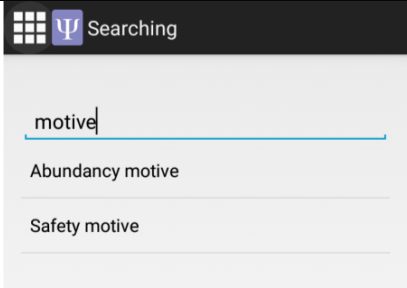
### 4.1. System Implementation

Implementation of the system is a step that must be done in the software development process. After passing through the analysis process by considering the possibility that will be implemented into the system so the next step is to implement what has been analyzed before. The system is built using Android Studio 1.2.2. In this system there are three menus: Main Menu, Psychology Terms Menu, About Menu.
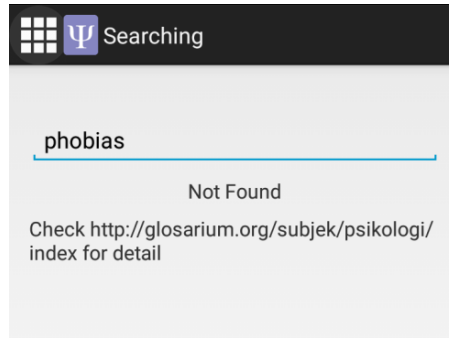
### 4.2. System Testing

System testing is performed to determine how the system work in terms of psychology searching using Horspool algorithm. In testing, which will looking for is a psychological term as its field. All search results are displayed according to the input that match on a search. In Table 4 on the next page can be seen terms data search result using Horspool

Table 4: The Results of the data terms search Horspool algorithm

| Name | Searching Result | Image Result |
| --- | --- | --- |
| A01 | Match |  |
| A02 | Match |  |
| A03 | Match |  |
| A04 | Match |  |

| A05 | Mismatch |
|-----|----------|



In Horspool algorithm, stored information (bad-match character) that is used to do shift when a mismatch occurs, and not just one character so accelerate the matching string. These algorithms do the matching from right to left.

Some of the search results above, input a few keywords that are used to test the system. For images A01, inputted keyword "motive". The search yielded the word "abundancy motive" and "Safety motive". As has been explained previously that the Horspool algorithm match the characters from right to left.

## 5. Conclusions

Based on the results of the Implementation of Horspool Algorithm In Making Dictionary of Psychology Terms on Android platform, it can be concluded that:

- Application of a string matching can be done using a variety of string matching algorithms, one of which is Horspool algorithm.
- Horspool algorithm proved able to resolve problems finding the degree of similarity of character.
- Results of testing for compatibility code indicates that the Horspool algorithm do the matching characters from right to left.

## References

[1]Baeza-Yates, R.A. & Regnier, M. (1992). Average running time of the boyer-moore-horspool algorithm. Journal Theoretical Computer Science 92 (1), 19-31.

[2]Charras, C. & Lecroq, T. (1997). Handbook of Exact String Matching Algorithms. Oxford University Press: United Kingdom.

[3]Cormen, T.H., Leiserson, C.E., Rivest, R.L. & Stein, C. (1994). Introduction to Algorithms. McGraw-Hill Book Company: North America.

[4]Effendi, D., Hartono, T. & Kurnaedi, A. (2013). Penerapan string matching menggunakan algoritma

[5]Boyer-Moore pada translator bahasa Pascal ke C. Majalah Ilmiah Unikom 11 (2), 262-271.

[6]Hasugian, J. (2006). Penggunaan bahasa alamiah dan kosa kata terkendali dalam sistem temu balik informasi berbasis teks. Jurnal Studi Perpustakaan dan Informasi 2 (2), 72-80.

[7]Horspool, R.N. 1980. Practical fast searching in strings. Journal Software Practice and Experience 10 (), 501-506.

[8]nuth D.E., Morris (Jr) J.H. & Pratt V.R. (1977). Fast pattern matching in strings. SIAM Journal on Computing 6 (1), 323-350.

[9]Parno, Dharmayanti & Rahmansyah, N. (2011). Aplikasi mobile kamus istilah psikologi berbasis Android 2.2. Prosiding Psikologi, Ekonomi, Sastra, Arsitektur & Sipil (PESAT), pp 121-128.

[10]Saptari, J. & Purwono. (2006). Temu kembali informasi bibliografi dengan bahasa alami pada field judul dan subjek : studi efektifitas katalog induk terpasang perpustakaan UGM. Jurnal Berkala Ilmu Perpustakaan dan Informasi 3 (1), 1-13.

[11]Sheik, S.S., Aggarwal, S.K., Poddar, A., Balakhrisnan, N. & Sekar, K. (2004). A fast pattern matching algorithm. Journal of Chemical Information and Computer Science 44 (4), 1251-1256.

[12]Singh, R. & Verma, H.N. (2011). A fast string matching algorithm. International Journal of Computer Technology and Applications 2 (6), 1877-1883.

[13]Wibowo, A. (2012). Peningkatan performansi sistem temu balik informasi dengan metode phrasal translation dan query expansion. Skripsi. Politeknik Negeri Batam.